

糖尿病患罹患十大癌症之風險預測系統

組員名單：莊佳怡、謝冠儀、郭采羚、王家修

指導教授：陳光琦

糖尿病 (diabetes mellitus, DM) 是一種慢性代謝疾病，特徵是患者的血糖長期高於標準值。根據國民健康署 106-109 年「國民營養健康狀況變遷調查」，18歲以上國人糖尿病盛行率為11.1% (男性 12.4%；女性 9.7%)，預估全國約有218.6萬人罹患糖尿病。

癌症 (cancer) 又名惡性腫瘤 (malignant tumor)，是控制細胞分裂增殖機制失常而引起的人體細胞疾病。惡性腫瘤生長快速，會侵犯身體重要部位，甚至經由體內循環或淋巴系統轉移到身體其他部位，對人體危害極大。研究顯示糖尿病與癌症高度關聯，糖尿病患的罹癌風險或死於癌症風險都比非糖尿病患高，雖然目前癌症無法根治，但若能早期發現並進行治療，就能大幅提高病患存活率。因此，本專題將針對糖尿病患建構罹癌風險之預測模型。

本專題將從臺灣全民健保資料庫擷取研究對象之病歷資料，進行預處理、整合，然後將資料集依比例切割成訓練集 (train set)、驗證集 (validation set) 及測試集 (test set)。再以機器學習 (machine learning) 及深度學習 (deep learning) 方法提取重要特徵 (feature)、調整超參數 (hyper-parameter)，然後以不同的類神經網路 (neural network) 建立模型。

最後考量不同模型的正確性 (accuracy)、敏感性 (sensitivity，又稱召回率 recall)、特异性 (specificity)、精確性 (precision)、F1分數 (F1 score)、ROC曲線下面積 (area under the receiver operating characteristic curve, AUC or AUROC)、正確性及損失 (loss) 函式趨勢圖，選出最佳預測模型。

研究成果將以互動式網頁呈現風險預測之結果，並參考醫師的專業意見以及病患回饋，改進網頁。期望能在早期發現癌症，盡早就醫以增加糖尿病患癌症治癒之機率。

關鍵字：糖尿病、癌症、電子健康紀錄、深度學習、風險預測